molecular systems biology

Single-cell analysis of the population context advances RNAi screening at multiple levels

Berend Snijder, Raphael Sacher, Pauli Ramö, Prisca Liberali, Karin Mench, Nina Wolfrum, Laura Burleigh, Cameron C Scott, Monique H. Verheije, Jason Mercer, Stefan Moese, Thomas Heger, Kristina Theusner, Andreas Jurgeit, David Lamparter, Giuseppe Balistreri, Mario Schelhaas, Cornelis A. M. De Haan, Varpu Marjomäki, Timo Hyypiä, Peter J. M. Rottier, Beate Sodeik, Mark Marsh, Jean Gruenberg, Ali Amara, Urs Greber, Ari Helenius, Lucas Pelkmans

Corresponding author: Lucas Pelkmans, University of Zurich

Review timeline:

Submission date:	30 July 2011
Editorial Decision:	20 September 2011
Revision received:	26 January 2012
Editorial Decision:	28 February 2012
Revision received:	09 March 2012
Accepted:	09 March 2012

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 20 September 2011

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the two referees who accepted to evaluate your manuscript. As you will see from the reports below, referee #1 is supportive whereas referee #2 is much more reserved. We agree with reviewer #1 that the dissection of the impact of cell population heterogeneity on the outcome of RNAi phenotypic screens is a topic of potential interest to the readers of our journal. Referee #2 feels that the work "is of value to the RNAi screening community" albeit of somewhat limited scope due to the focus on image-based phenotypic assays. On balance, considering these evaluations and in view of the increasing number of published image-based RNAi phenotypic screens, we feel that we can consider a major revision of the manuscript.

One of the major points raised by the two reviewers refers to the need of a less biased presentation of the results and avoid over-generalization. Thus, the study could be presented in a more neutral way as an extensive test of the potential impact of population heterogeneity and report the results in a much more balanced way (reviewer #2: "more forthcoming about what the majority of their data shows regarding their hypotheses"). Accordingly, rather than presenting context-based normalization as general tool, one of the major outcome of this analysis could be to provide a more in-depth discussion of which situations/biological processes are expected to require context-based normalization and in which situations this correction does not appear to explain the observed lack of correlations between screens/assays. Along the same lines, it would be important to analyze of how

simplified corrections, for examples only based on cell number, would alter the technical and biological conclusions. Finally, in terms of biological interpretation, while the fraction of changed hit ranks is discussed, it would be important to analyze more systematically the impact at the pathway level or in terms of functional enrichment that would not have been detectable/significant without the context-based correction.

We appreciate that the full description of the methods may not fit within the main paper. Please include nevertheless a Materials and Method section in the main paper and refer to Supplementary information for the full description of the methods. Please include also in the Materials and Method section of the main text a section on "data availability". The websiste www.infectome.org does not appear to be functional yet. According to the policy of our journal (http://www.nature.com/msb/authors), it is important that, for the 34 small-scales and 7 large scale screen used in this study, the images, the corresponding extracted features and the infection data are made publicly available in a form that allows other to repeat your analysis or to build upon and integrate your data in other studies.

*** PLEASE NOTE *** As part of the EMBO Publications transparent editorial process initiative (see our Editorial at http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology will publish online a Review Process File to accompany accepted manuscripts. When preparing your letter of response, please be aware that in the event of acceptance, your cover letter/point-by-point document will be included as part of this File, which will be available to the scientific community. More information about this initiative is available in our Instructions to Authors. If you have any questions about this initiative, please contact the editorial office msb@embo.org.

If you feel you can satisfactorily deal with these points and those listed by the referees, you may wish to submit a revised version of your manuscript. Please attach a covering letter giving details of the way in which you have handled each of the points raised by the referees. A revised manuscript will be once again subject to review and you probably understand that we can give you no guarantee at this stage that the eventual outcome will be favourable.

Yours sincerely,

Editor
Molecular Systems Biology

http://www.nature.com/msb

Referee reports:

Reviewer #1 (Remarks to the Author):

Here Snijder et al. describe an elegant method in order to account for population context in RNAi screens for regulators of virus uptake. There is a timeliness to this work as a number a recent RNAi screens, especially for genes involved in infection, have show strikingly poor correlation. Moreover, the question of how do deal with population heterogeneity is an important issue given that we now have methods to rapidly quantify the phenotype of single cells in very high-throughput.

Overall I find this is a thorough technical description of a powerful means by which to consider population context. I am not sure that this work itself provides a great deal of insight into the systems regulating virus uptake, but I think the readers of MSB will find this a paper of interest. I recommend the manuscript for publication assuming some revisions can be made, especially which improve the overall clarity and precision of their language.

Major Points:

1) In summary the authors claim that RNAi screens are "globally improved" when population context is taken into account.

I have a few comments regarding this statement:

- This improvement is observed when compared to screens that did not consider population heterogeneity (e.g. such the initial infection screens), and/or differential effects of different siRNAs. There is certainly growing appreciation of both these factors (which the authors themselves have described in the past), and many of the techniques used to analyze initial screens either did not, or could not, consider them. Or to put it another way, the screening methods to which this study makes comparison should now likely be considered suboptimal, and thus I do not find the observation that these screens could be improved upon is itself particularly surprising. With this said I do find the method an extremely attractive means by which to address these important factors.

- Clearly there is an overall improvement in the correlation between siRNA phenotypes, but the improvement in the correlation between screens using the same virus seems relatively mild (assuming I am reading Fig. 5d correctly).

This result is quite important as it suggests that of the poor reproducibility that has been observed in previous screens is not entirely driven by population context.

- Similarly some discussion of when consider population context is may be a contributing factor, or is not, is warranted here. Moreover, it is important to note that some other methods (e.g. Collinet et al. 2009) have achieved similar results regarding siRNA reproducibility without (at least directly) considering context.

- That population context improves the "interpretability" of screens due to the fact that splicing factors are no longer enriched in one screen is a very qualitative judgmental statement. I do not feel this adds anything to the manuscript.

2) At times, this manuscript is very difficult to go through, and perhaps overly technical for a general journal such as MSB. What the authors are sometimes analyzing/quantifying is sometimes very difficult to decipher. The readability could also be improved by reworking some of the figures and figure legends such that it more clear exactly what types of analysis are being performed. In some cases simple examples might help. For example, small "workflows" or descriptions of the methods being used to generate different results may help more so than the some of the figure panels that, while colorful, are not particularly informative.

3) The authors often implement analysis methods that quantify siRNAs scoring as hits. I would suggest the inclusion a clear figure that shows how the consistency of the phenotype, regardless of whether the gene is a hit, that is elicited by different siRNAs is improved when population context is taken into account.

Minor comments:

- It is worth noting that all the screens have presumably been performed by the same group and thus a great deal of potential factors which may contribute to poor reproducibility in screens performed between different labs has been completely eliminated. It will be interesting to see if these methods can be applied to different datasets. But I do not suggest this is necessary here.

- I find the observation that levels of Trio and Abl protein expression are predictive of relative changes in infection following RNAi a very interesting result. In fact this suggests that heterogeneous levels of protein expression are a significant contributing factor to virus uptake. That protein levels correlate with cell position in culture is (to me) quite unexpected.

In fact it may be that many of the effects the authors have observed here and in previous work are driven by protein heterogeneity.

I would prefer this is discussed more at length.

Reviewer #2 (Remarks to the Author):

In the manuscript, "Single-cell analysis of the population context globally improves large-scale RNAi screens" by Snijder et al., the authors attempt to address the variances in phenotype that are caused in RNAi screens by siRNAs altering cell context. The main hypothesis of the paper is that there are direct effects of siRNAs that perturb the phenotype of interest due to alterations in the levels of proteins that impact the phenotype. Alternatively, there are siRNAs that indirectly alter the phenotype because they alter cell number and thus cell context. The authors previously presented both research articles and review articles investigating and discussing that tissue culture cells alter some of their physiology depending on their density and context. These context dependent alterations appear to increase with increasing cell number as this also will alter the ratio of edge cells to interior cells, as the cells in the well become more crowded and less of them occupy an edge, and more are in an interior environment. Once the well in confluent then all the cells will be on the interior and the edge phenotype would likely disappear altogether. The authors have previously shown that virus replication differs in edge cells versus interior cells. Therefore siRNAs that impact either cell number, the ratio of cells on an edge vs. an interior, or the processes that are modulated by these contextual forces, will all potentially alter infection rates. By correcting for these "indirect effects" the authors state that they increase the percentage overlap between screens as well as the number of siRNAs that represent direct targets. The authors present evidence that these corrections also improve the veracity of siRNA targets in screens addressing phenotypes other than viral replication.

While RNAi screens are producing transforming discoveries monthly, there are well appreciated significant drawbacks to this approach, i.e. off target effects, false negatives, and toxicity. These pitfalls are not insurmountable, but do require a very high degree of validation to convincingly demonstrate the worth of RNAi screening data. In addition, major attention has been focused on the lack of concordance between screens, the HIV screens in particular, in large part due to a lack of insight in follow-up efforts and reviews. These works focused on exact gene overlaps, and not on enrichment in pathways and complexes, an obvious point addressed in the paper by Bushman et al. Therefore this attempt to improve screen interpretation, better understand screen discordance and enhance the integration of large datasets is of value to the RNAi screening community, yet its primary applicability would be restricted to those using image-based approaches focused on phenotypes that are impacted by cell context.

Concerns:
Much of the text and labels of figures in the supplemental was most unfortunately garbled or missing letters.

In the introduction, the authors state that wide perturbations obtained with different siRNAs targeting the same gene have suggested wide-spread off target effects, casting doubt on the usefulness of the approach. The authors should please acknowledgement that off-target effects are dependent on the siRNAs or siRNA library being used with the earlier reagents, or those currently of poorer quality, not avoiding micro-RNA sequences and also lacking later modifications, i.e. antisense and sense strand methylation that now greatly minimize OTEs.

Also this suggests that the author's current methods will correct for off target effects, something which is not fully assessed in this paper and should be stated, that is we do not know what percentage of the cell context dependent effects are on targets or off targets with the given level of data and validation provided.

The author's argue that indirect effects that alter cell number and context produce major alterations in hit selection in RNAi screens. The authors address this phenomena using a matlab based approach that is partially outlined in the methodology but relies heavily on software downloads. Indeed the authors refer the reader to the Pelkman lab's website for the use of matlab downloads. However if the authors wish this approach to be generally used, as say a Snijder-Pelkmans correction, then they should add a more user-friendly version of the analysis so that past and present screeners can use this improvement, starting with analyzing their stored image tiffs. This should be placed on the

website as a resource. Otherwise a correction for cell number will most likely be the end-result of this work.

The authors should comment on how image-based screen represents a significant advantage over plate based screening strategies because one needs images to assess the impact on cell colony morphology.

The authors should comment on why there is still such a low rate of concordance between similar screens that differ only in the strain of HeLa cells used even after their correction is applied. Are these really all HeLa cells? Did they receive these lines from the ATCC or other tissue culture cell bank source? If cell number alterations are on target indirect effects, then the same siRNA should produce similar alterations in cell number/context across HeLa cell strains. If the indirect effects are on target shouldn't they match across screens for the same virus? Did the authors see this was the case?

The authors should please refer to the library they use as a druggable genome library or a sub-genomic library, genome-scale is not appropriate since that would involve > 3 times as many genes as screened.

Page 5 line 26: Figure 14 legend states that 50% of the indirect effects can be accounted for by cell number, therefore the authors comment "which cannot be inferred from knowing the total number of cells" should be changed "to which can partially (50%) be inferred from knowing the total number of cells."

Page 6 line 26: The authors state "This demonstrates that a large part of the changes in infection in these RNAi screens are determined by siRNA-induced changes in cell population context." What percent of hits per screen is attributable to indirect effects for each screen? This should be provided in a table. Can this be corrected primarily by accounting for cell number since kinases will impact the cell number parameter substantially? How strong are these effects versus direct effects? Can we preferentially filter a significant percentage of indirect effects by looking at their relative strength compared to direct effects? i.e. increase the cut-off for hits selection. The authors mention both the growth-promoting first siRNA for DYRK3 and the growth-inhibitory effects of AK5, as both working indirectly via alterations in cell number, doesn't this strongly suggest that cell number alone can be used as a filter for indirect effects. Why use a complex analysis program that likely the majority of screeners will not be able to master if most of the effect can be controlled for or spotted using simpler approaches? The authors should please address this very practical implementation concern.

Page 7 line 15 Figure S30 is mentioned as providing validation of direct effects. The effects shown with many of the siRNAs are moderate, in some cases less than two fold change. Therefore corresponding expression levels of the targeted proteins or mRNAs are needed to help accurately interpret if these effects are on target, either westerns or q-PCR to show that the level of knockdown matches the phenotype. Additionally data should be shown for both the shRNA-mediated knockdown and the restored expression of AAK1. It is also problematic to rescue a knockdown by simply overexpressing a cDNA; it must be resistant to the expressed shRNA to demonstrate specificity. Is the shRNA against the 3'UTR of AAK1? Adequate information for this figure should be provided so that the results can be interpreted and reproduced, i.e. siRNA and shRNA sequences and methods. Also what do the error bars represent? How many times was this experiment performed?

Page 8 lines 13-17, the authors state that for the minority of the large screens, 3/7, masked or indirect effects accounted for 25-50% of all siRNA phenotypes and refer to Fig. 28 which also shows that for 4/7 screens the impact of masked and indirect was negligible. The authors should be less biased and more forthcoming about what the majority of their data shows regarding their hypotheses. Why is there such a stark difference in how the viruses behave in regard to the cell context? It is hard with the garbled figure legends to tell what virus is shown for what graph, this should be corrected of course. However it suggests that cell context is NOT important for the majority of the larger screens that the authors have completed, suggesting that it may be indirect modulations of discrete host cells pathways, i.e. endocytosis, that impact certain viruses more than others, or certain cell lines more than others. Some viruses use more than one host cells endocytic

pathway to infect, does reliance on CME result in a more significant cell context dependent phenotype? Again this argues that much of what the authors are discussing may be attributable to the cell context modulation of a few cellular actions, i.e. lipid regulation and CME for example, two pathways that were previously reported by the authors as being modulated by cell context. Therefore these large data sets suggest that if the phenotypic readout for a screen involves a cell context sensitive pathway then you should plot infection (or whatever your readout of interest is) against cell number. If there is a correlation observed then you might filter out "higher cell number" hits as a correction, or test the hits with different densities of cells per well in follow up assays. Also since the kinase screens may have had more indirect effects per the authors; it suggests that they are not representative of the actions of siRNAs from a broader sampling of genes. This should be commented upon. Did the same viruses that had virtually no indirect effects in the larger screens demonstrate a different level of susceptibility to indirect effects in their respective kinase screens? And did this correlate with cell numbers?

The authors must publish or make available the percent infection data for ALL of the screens analyzed in this work, i.e. cell numbers, so that their results can be assesses appropriately by others. They should also include the relevant imaging analyses parameters they have calculated, so that the conclusions and methods can be assessed. This is very important since they are not focusing on specific siRNA hits but instead are strongly emphasizing the broad applicability of their methods and conclusions to the overarching data analysis and hit selection of entire screens to be used as a guiding resource for the RNAi screening community.

Why haven't cell context dependent events been observed or reported in image based small molecule screens for anti-viral compounds? One might imagine that the time course of the screen and the larger effects used for hit selection may have served to minimize the impact of compounds that increase or decrease cell growth. The authors should please comment on this and also provide cell number data for their validation experiments using GSK-626616 since this is not provided in Fig 30 panels g and h.

Page 10 line 16, the authors should note that the 7% overlap is of exact genes, there was significantly more overlap in pathways and complexes as the Bushman paper made more clear. In those instances different cell lines, viral preps, protocols and siRNA libraries were used. In the case of this work all of those were identical with the exception of the cell lines, which vary. Therefore the relevance of these observations to the comparison of screens between groups using much different approaches may be of less value. The authors should acknowledge the overlap in pathways and complexes in viral screen meta-analyses, as well as to the much greater homogeneity in their current screening efforts than those done by distinct groups.

Page 10 bottom paragraph and figure 5d; it appears that only three of 15 screens improved in assay correlation with correction, suggesting that cell context variability occurs in the minority of screens presented. Are these the same viruses that showed the most indirect effects in Fig S28? It is hard to tell because of the garbled legends. Regardless, the authors should please comment on the general relevance of their correction if the majority of their screens did not improve in correlation after correction.

1st Revision - authors' response 26 January 2012

Response to editor's points:

As you will see from the reports below, referee #1 is supportive whereas referee #2 is much more reserved. We agree with reviewer #1 that the dissection of the impact of cell population heterogeneity on the outcome of RNAi phenotypic screens is a topic of potential interest to the readers of our journal. Referee #2 feels that the work "is of value to the RNAi screening community" albeit of somewhat limited scope due to the focus on image-based phenotypic assays.

We agree that the full effects can only be addressed in image-based assays with single-cell

resolution, but the effects in principle apply to any RNAi screen that uses perturbed populations of cells. Ignoring the problem with averaged approaches will of course not make it go away. In some ways our work is thus particularly important for those applying non-image based screening methods to make them aware of the pitfalls in their approach.

Unfortunately, there is currently no comprehensive list of which cellular activities depend on the cell population context and it is therefore unclear to which extent our findings apply generally. However, the results in this manuscript (as well as our previously published and currently unpublished observations) now suggest that it may apply to a wide variety of cellular activities. Also, by now sharing easy-to-use modules (in MatLab and CellProfiler) that enable anyone to apply our methods we expect that the scope will be less limited.

On balance, considering these evaluations and in view of the increasing number of published image-based RNAi phenotypic screens, we feel that we can consider a major revision of the manuscript.

One of the major points raised by the two reviewers refers to the need of a less biased presentation of the results and avoid over-generalization. Thus, the study could be presented in a more neutral way as an extensive test of the potential impact of population heterogeneity and report the results in a much more balanced way (reviewer #2: "more forthcoming about what the majority of their data shows regarding their hypotheses").

In response to this point we have undertaken the following steps:

- We extended and generalized the set of screens analyzed by adding two genome-wide screens for cellular cholesterol levels as well as for cell size. These screens further generalize our findings beyond virus infection, and were performed in another lab and city and different screening facility and with different siRNA libraries and a different cell line.
- We have included an excel file (Supplementary Table 7) giving an overview of hit-list overlap improvements, showing improvements across a wide array of assays
- We have rewritten the manuscript to emphasize the generality of our findings and methods where appropriate and toned down our generalization where not.
- Wherever applicable we have expanded on the reporting of the majority of our data, as well as enabled full accessibility to the raw data, in order to address the feeling that we are not forthcoming about our results.

Accordingly, rather than presenting context-based normalization as general tool, one of the major outcome of this analysis could be to provide a more in-depth discussion of which situations/biological processes are expected to require context-based normalization and in which situations this correction does not appear to explain the observed lack of correlations between screens/assays.

We have added a paragraph to the discussion on this topic.

Along the same lines, it would be important to analyze of how simplified corrections, for examples only based on cell number, would alter the technical and biological conclusions.

We have now included a comprehensive analysis comparing two simplified cell-number corrections (lowess normalization and a cutoff) with our approach (see Supplementary Figure 15). This reveals that both these simplified methods are not sufficient to correct for population context-mediated effects. Lowess (the more advanced method of the two) for instance only corrects 81 of the 203 genes in the same way as our method (for virus infection, on average over all 7 DG screens), and introduces genes as a result of overfitting – i.e. without having a biologically motivated reason. As expected, the cutoff method performs even worse.

Finally, in terms of biological interpretation, while the fraction of changed hit ranks is discussed, it would be important to analyze more systematically the impact at the pathway level or in terms of functional enrichment that would not have been detectable/significant without the context-based correction.

In response to this point, we have now included several meta-analyses of our screens that fall in two categories:

1) At the level of functional annotation overlap improvement. Here we used cell size as readout common to all 7 druggable genome screens. As expected by the increased hit list overlap, annotation-overlap is improved even more between all 7 druggable genome screens. We now give Supplementary Table 5 with full annotation enrichment results for these screens, as well as an excel file with all cell size readouts (Supplementary Table 6).
2) At a systems-level interpretation of the clustering of 34 small-scale RNAi screens of virus infection. This shows the strength of cluster-analysis at the meta-level using population context-corrected readouts of parallel RNAi screens of virus infection (See Fig. 7D).

We appreciate that the full description of the methods may not fit within the main paper. Please include nevertheless a Materials and Method section in the main paper and refer to Supplementary information for the full description of the methods.

We have now included a Materials and Methods section to the main manuscript, and refer to the supplementary information for additional details.

Please include also in the Materials and Method section of the main text a section on "data availability". The webisite www.infectome.org does not appear to be functional yet. According to the policy of our journal (http://www.nature.com/msb/authors), it is important that, for the 34 small-scales and 7 large scale screen used in this study, the images, the corresponding extracted features and the infection data are made publicly available in a form that allows other to repeat your analysis or to build upon and integrate your data in other studies.

We have now included this in the Materials and Methods section, and have undertaken the following:

- All data of the 34 small-scale RNAi screens of the 17 virus infections are fully accessible at www.infectome.org. This includes images (originally 800GB of image data) and raw measurements, that are browsable and searchable on the website.
- The entire dataset of the SV40 infection druggable genome screen is available as a supplementary Excel file (Supplementary Table 8). Image data will follow on the infectome.org website, but we cannot guarantee its availability for this resubmission due to the technical challenges associated with the publication of such a large dataset in its entirety.
- Entire cell size readouts and full population context parameters of all 7 druggable genome screens will be published with this study as Excel file.
- The infection data of the Vaccinia virus druggable genome screen will be published independently (including extensive cell biological follow-up work) and the data will be released on infectome.org in a timely manner.
- All other virus screen data will follow on www.infectome.org
- The cellular cholesterol level screens (1 druggable genome and 2 genome-wide) will be published elsewhere.

Reviewer #1 (Remarks to the Author):

Here Snijder et al. describe an elegant method in order to account for population context in RNAi screens for regulators of virus uptake. There is a timeliness to this work as a number a recent RNAi screens, especially for genes involved in infection, have show strikingly poor correlation. Moreover, the question of how do deal with population heterogeneity is an important issue given that we now have methods to rapidly quantify the phenotype of single cells in very high-throughput.

Overall I find this is a thorough technical description of a powerful means by which to consider population context. I am not sure that this work itself provides a great deal of insight into the systems regulating virus uptake, but I think the readers of MSB will find this a paper of interest. I recommend the manuscript for publication assuming some revisions can be made, especially which improve the overall clarity and precision of their language.

Major Points:

1) In summary the authors claim that RNAi screens are "globally improved" when population context is taken into account.

I have a few comments regarding this statement:

- This improvement is observed when compared to screens that did not consider population heterogeneity (e.g. such the initial infection screens), and/or differential effects of different siRNAs. There is certainly growing appreciation of both these factors (which the authors themselves have described in the past), and many of the techniques used to analyze initial screens either did not, or could not, consider them. Or to put it another way, the screening methods to which this study makes comparison should now likely be considered suboptimal, and thus I do not find the observation that these screens could be improved upon is itself particularly surprising. With this said I do find the method an extremely attractive means by which to address these important factors.

- Clearly there is an overall improvement in the correlation between siRNA phenotypes, but the improvement in the correlation between screens using the same virus seems relatively mild (assuming I am reading Fig. 5d correctly).

We understand the reviewer's concern. Calculating the overall correlation between all data-points of two screens is just one measure of similarity, which is strongly determined by the bulk of the data-points. These are typically non-hits. This is exactly why we also included hit list overlap improvements in our first submission. To further clarify the difference between correlation and hit list overlap we have now further quantified improvements in hit list overlap for the 34 small-scale RNAi screens (see Figure 4A and 4D (was Fig. 5)). This shows improvements for 9 of the 15 viruses (while 4 stay equal and 2 decrease).
Furthermore, we analyzed hit list overlap of additional screens, all showing improvements in hit list overlap.

This result is quite important as it suggests that of the poor reproducibility that has been observed in previous screens is not entirely driven by population context.

We completely agree with the reviewer that not all problems in reproducibility are due to population context-mediated effects. Different cell lines can have truly different pathways, which will not be corrected by population context correction. Many examples exist of cell line-specific molecular regulation of cellular activities.
Our method can only make screening results from different cell lines more comparable if the activity screened is different in its population context dependencies. Besides this, technical and experimental variations will always introduce irreproducible noise in the data. We also make this point in the discussion.

- Similarly some discussion of when consider population context is may be a contributing factor, or is not, is warranted here.

We have now included a paragraph in the discussion on this topic, discussing when population context may be an important factor to look at. We further refer to our recent opinion article in NRMCB for a more detailed discussion.

Moreover, it is important to note that some other methods (e.g. Collinet et al. 2009) have achieved similar results regarding siRNA reproducibility without (at least directly) considering context.

In Collinet et al. there is no detailed comparison made between the reproducibility of their method and any other method (for instance a univariate method), so they cannot make any claims for **improvement** in siRNA or screen reproducibility. Of course, using 7 individual siRNAs per gene will more likely give 2 siRNAs with similar phenotype even when measurements of a limited number of cells are used and are averaged over all cells. This 'apparent' consistency just reflects the use of brute force, which is unnecessary if one would use the approach presented here. Thus, also in the Collinet paper we expect that our approach will improve siRNA consistency. In fact, the poor

reproducibility of their (population-averaged) siRNA phenotypes (as mentioned in that paper) is consistent with our observation of poor reproducibility observed when not accounting for population context. We cite their work accordingly.

- That population context improves the "interpretability" of screens due to the fact that splicing factors are no longer enriched in one screen is a very qualitative judgmental statement. I do not feel this adds anything to the manuscript.

We agree with this point and have removed the comment on splicing factors.

2) At times, this manuscript is very difficult to go through, and perhaps overly technical for a general journal such as MSB. What the authors are sometimes analyzing/quantifying is sometimes very difficult to decipher. The readability could also be improved by reworking some of the figures and figure legends such that it more clear exactly what types of analysis are being performed. In some cases simple examples might help. For example, small "workflows" or descriptions of the methods being used to generate different results may help more so than the some of the figure panels that, while colorful, are not particularly informative.

We apologize for this and have extensively rewritten the text to improve the readability and made figure and legends more clear. We believe it has improved, although we cannot completely avoid some technical language. At certain points this is necessary in order to be entirely specific and to avoid confusion.

3) The authors often implement analysis methods that quantify siRNAs scoring as hits. I would suggest the inclusion a clear figure that shows how the consistency of the phenotype, regardless of whether the gene is a hit, that is elicited by different siRNAs is improved when population context is taken into account.

We agree that we have not explicitly mentioned this. When our method increases 2/3 and 3/3 hits (genes for which 2 or 3 out of 3 siRNAs give a significant and similar direct perturbation), it by definition reduces 1/3 non-hits, and increases 0/3 non-hits. In other words, consistency improves regardless of whether a gene is a hit. We have made this now more explicit in Supplementary Figure 31.

Minor comments:

- It is worth noting that all the screens have presumably been performed by the same group and thus a great deal of potential factors which may contribute to poor reproducibility in screens performed between different labs has been completely eliminated. It will be interesting to see if these methods can be applied to different datasets. But I do not suggest this is necessary here.

We now show that our methods can be applied to different datasets, as we have included the analysis of two genome-wide RNAi screens performed in a different laboratory (in Geneva/Lausanne), with a different readout (Filipin staining), different siRNA library (for the genome-wide screens) and different protocols. We are happy to report that our approach also improved hit list overlap of the Filipin readout for the two genome-wide screens.
Furthermore, by analyzing the common cell size readout between our 7 druggable genome screens and their screens, we show that our approach improves hit list overlap also between screens performed at different locations and screening facilities.

- I find the observation that levels of Trio and Abl protein expression are predictive of relative changes in infection following RNAi a very interesting result. In fact this suggests that heterogeneous levels of protein expression are a significant contributing factor to virus uptake. That protein levels correlate with cell position in culture is (to me) quite unexpected.

We have added a sentence referring to other work that has made the observation that protein levels can vary between single cells of a monoclonal cell population depending on cell size (in yeast, using FACS) and cell density (in mammalian cells, using population averaged micro-arrays).

In fact it may be that many of the effects the authors have observed here and in previous work are

driven by protein heterogeneity.

I would prefer this is discussed more at length.

We agree, and have done so. We in addition expect that heterogeneity in protein activitiy or modification (eg. phosphorylation), subcellular localization, etc., initiated by signaling networks that sense the microenvironment, will contribute to the observed effects. We also discuss this further in our recent NRMCB opinion article.

Reviewer #2 (Remarks to the Author):

In the manuscript, "Single-cell analysis of the population context globally improves large-scale RNAi screens" by Snijder et al., the authors attempt to address the variances in phenotype that are caused in RNAi screens by siRNAs altering cell context. The main hypothesis of the paper is that there are direct effects of siRNAs that perturb the phenotype of interest due to alterations in the levels of proteins that impact the phenotype. Alternatively, there are siRNAs that indirectly alter the phenotype because they alter cell number and thus cell context.

While changes in cell number usually lead to changes in population context, we show in Fig 2 that cell population context can equally change upon a perturbation without that this is reflected in a change in cell number.

The authors previously presented both research articles and review articles investigating and discussing that tissue culture cells alter some of their physiology depending on their density and context.
These context dependent alterations appear to increase with increasing cell number as this also will alter the ratio of edge cells to interior cells, as the cells in the well become more crowded and less of them occupy an edge, and more are in an interior environment. Once the well in confluent then all the cells will be on the interior and the edge phenotype would likely disappear altogether.

The authors have previously shown that virus replication differs in edge cells versus interior cells.

…and also depending on local cell density, and on cell size.

Therefore siRNAs that impact either cell number, the ratio of cells on an edge vs. an interior, or the processes that are modulated by these contextual forces, will all potentially alter infection rates. By correcting for these "indirect effects" the authors state that they increase the percentage overlap between screens as well as the number of siRNAs that represent direct targets. The authors present evidence that these corrections also improve the veracity of siRNA targets in screens addressing phenotypes other than viral replication.

While RNAi screens are producing transforming discoveries monthly, there are well appreciated significant drawbacks to this approach, i.e. off target effects, false negatives, and toxicity. These pitfalls are not insurmountable, but do require a very high degree of validation to convincingly demonstrate the worth of RNAi screening data.

We here show that part of these drawbacks (false negatives/positives, toxicity, and inconsistency between siRNAs targeting the same gene) can partially be tackled by not averaging single-cell readouts and correcting for population context-mediated effects.

Furthermore, this view on RNAi screens has limited the use of full RNAi screening datasets for the systems-level interpretation of such datasets. Obviously, to be sure of RNAi screening hits a full validation of individual genes is required. But for use in systems approaches we here show that improving the overall quality of RNAi screening datasets is required for a meaningful systems level interpretation. It is exactly this use of RNAi screening data that we feel is of interest to the readership of MSB.

In addition, major attention has been focused on the lack of concordance between screens, the HIV screens in particular, in large part due to a lack of insight in follow-up efforts and reviews. These works focused on exact gene overlaps, and not on enrichment in pathways and complexes, an

obvious point addressed in the paper by Bushman et al.

We now also show that our approach strongly improves the overlap of enriched functional annotations (i.e. pathways and complexes) between replicate large-scale screens (as expected by the increase in exact hit overlap). See Supplementary Figure 30 and further discussion below.

Therefore this attempt to improve screen interpretation, better understand screen discordance and enhance the integration of large datasets is of value to the RNAi screening community, yet its primary applicability would be restricted to those using image-based approaches focused on phenotypes that are impacted by cell context.

We are happy that the reviewer sees value in our method, and we agree that our methods are only applicable to image-based readouts with single-cell resolution.

However, our findings are just as important for any RNAi screen that uses cell population-averaged readouts, be it western blot or plate-reader assays. We therefore believe our findings will make people more aware of these pitfalls of population averaging when interpreting perturbation phenotypes in general.

We further agree with the reviewer that our methods only apply to phenotypes impacted by the cell population context. Unfortunately, there is currently no overview of which cellular activities are impacted by the population context. Given the results published previously (Snijder et al, Nature, 2009), presented here (for 17 viruses, cellular cholesterol levels, cell size, and endocytosis) and unpublished observations, we expect and indeed see many cellular processes to be impacted by the cell population context.

Concerns:
Much of the text and labels of figures in the supplemental was most unfortunately garbled or missing letters.

We apologize for the inconvenience. Unfortunately this must have happened during the PDF conversion by the online submission tool of MSB. We would like to note however that the original PDF, which the reviewer could also have accessed from the MSB manuscript tracking system, seemed fine.

In the introduction, the authors state that wide perturbations obtained with different siRNAs targeting the same gene have suggested wide-spread off target effects, casting doubt on the usefulness of the approach. The authors should please acknowledgement that off-target effects are dependent on the siRNAs or siRNA library being used with the earlier reagents, or those currently of poorer quality, not avoiding micro-RNA sequences and also lacking later modifications, i.e. antisense and sense strand methylation that now greatly minimize OTEs.

We agree with the reviewer and have added a sentence acknowledging this.

Also this suggests that the author's current methods will correct for off target effects, something which is not fully assessed in this paper and should be stated, that is we do not know what percentage of the cell context dependent effects are on targets or off targets with the given level of data and validation provided.

We agree with the reviewer and have added a sentence in the discussion to this extent.

The author's argue that indirect effects that alter cell number and context produce major alterations in hit selection in RNAi screens. The authors address this phenomena using a matlab based approach that is partially outlined in the methodology but relies heavily on software downloads. Indeed the authors refer the reader to the Pelkman lab's website for the use of matlab downloads. However if the authors wish this approach to be generally used, as say a Snijder-Pelkmans correction, then they should add a more user-friendly version of the analysis so that past and present screeners can use this improvement, starting with analyzing their stored image tiffs.

This should be placed on the website as a resource. Otherwise a correction for cell number will

most likely be the end-result of this work.

We agree with the reviewer and have therefore created an implementation of our novel methods that is much more user-friendly. Specifically, we have created two modules (MeasurePopulationContext and PopulationContextCorrect) for CellProfiler (v1) that implement both our methods for the measurement of all features of the population context used in this study, as well as for the automated modeling and correction of the influence of the population context on any given cellular measurement. These modules are provided for download from the www.infectome.org website with an example dataset and pipeline, and extensive documentation (on the website, and in the CellProfiler help function) showing how to use these modules.

The authors should comment on how image-based screen represents a significant advantage over plate based screening strategies because one needs images to assess the impact on cell colony morphology.

We completely agree and we have added this to the discussion.

The authors should comment on why there is still such a low rate of concordance between similar screens that differ only in the strain of HeLa cells used even after their correction is applied.

We have added this to the discussion.

However, we don't quite agree that there is such a low degree of concordance between our screens. The small-scale screens performed with the same virus but on different strains of HeLa cells have a correlation coefficient of on average 0.6, (Many screens e.g. Collinet et al., 2010, only reach this correlation for replicates within the same screen). Furthermore, with regard to improvements in overlap, we have now significantly extended the analysis of this part, also on additional large-scale screens, showing that improvement in overlap is considerable.

Are these really all HeLa cells?

Except for the A431 cells used in this study all other cell lines are indeed HeLa cells.

Did they receive these lines from the ATCC or other tissue culture cell bank source?

The origin of these cell lines was described in the previous version of the Supplementary Materials and Methods already. These exact cell lines have been recently and extensively used in a number of large-scale RNAi screens:
HeLa MZ: Collinet et al, Nature, 2010.
HeLa KY: Mitocheck consortium: Neumann et al, Nature, 2010.
HeLa CNX: Cenix Biosciences (http://www.cenix.com/)
HeLa TDS: Ralph Kittler and Frank Bucholtz and the High-Throughput Technology Development Studio at the MPI in Dresden (http://www.mpi-cbg.de/facilities/profiles/ht-tds.html)

If cell number alterations are on target indirect effects, then the same siRNA should produce similar alterations in cell number/context across HeLa cell strains.

That is correct, assuming knockdown efficiency is the same. We see cell number effects to be reproducible across different HeLa cell strains. For the 34 small-scale screens, the median Pearson correlation coefficient over total cell number values between all assay-pairs of different HeLa strains is 0.46 (average of n = 359). However, how cell number determines other factors of the cell population context (Sup. Fig. 8), and how this in turn affects virus infection (Fig 1) can differ. As we show, even small changes in cell number and/or the population context can in turn have large effects on measurements of context-dependent cellular activities. Thus a slight difference in silencing efficiency, although being on-target, can be propagated and amplified in non-linear ways through the population context resulting in quite strong differences in cellular activity. Furthermore, it is important to keep in mind that siRNA effects can alter the population context significantly without changing cell numbers (Fig. 2).

If the indirect effects are on target shouldn't they match across screens for the same virus?

Did the authors see this was the case?

Yes, we partially see that: A substantial part of indirect effects are reproducible between different siRNAs in the same assay (corr. coef = 0.7, see Sup. Figure 9), and between different cell lines (see above, corr. coef = 0.46), and therefore likely on target. This is why accounting for indirect effects on average slightly reduces the correlation between infection in cells perturbed with different siRNAs targeting the same gene (shown in Suppl. Figure 9). However, the siRNA phenotypes of hits (as opposed to that of all siRNA phenotypes in the screen, as measured by correlation) become more consistent after accounting for these indirect effects on infection, as shown in Sup. Figs 12 and 31. This is furthermore consistent with the observation that a large part of variance in infection phenotypes can be attributed to indirect effects (as shown by the high similarity of measured and predicted infection levels in Sup. Figs. 13 and 14).

The authors should please refer to the library they use as a druggable genome library or a sub-genomic library, genome-scale is not appropriate since that would involve > 3 times as many genes as screened.

We have included 2 genome-wide siRNA screens confirming our claim that our results apply to genome-scale screens. We have further rephrased our nomenclature to “small”, “medium”, “large” (or “druggable genome”) and “genome-wide” screens. We feel that the screening of 21.000 siRNAs repeated three times (as we have done for each of the 7 druggable genome screens) does warrant the use of the word “large”.

Page 5 line 26: Figure 14 legend states that 50% of the indirect effects can be accounted for by cell number, therefore the authors comment "which cannot be inferred from knowing the total number of cells" should be changed "to which can partially (50%) be inferred from knowing the total number of cells."

We have changed the sentence accordingly. However, correcting just for cell number effects does not tell you which 50% of your corrections should have been applied. In other words, for each correction or lack of correction there is a 50% chance that it is applied inappropriately.

Page 6 line 26: The authors state "This demonstrates that a large part of the changes in infection in these RNAi screens are determined by siRNA-induced changes in cell population context." What percent of hits per screen is attributable to indirect effects for each screen? This should be provided in a table.

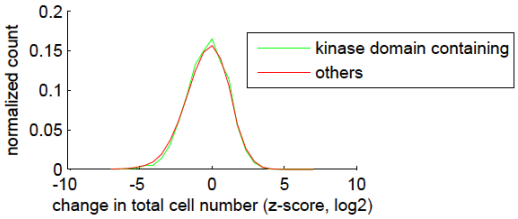
This was actually shown in Figure 5A (now relabeled to Figure 4A): The amount of hits changed upon accounting for indirect effects represents the amount of hits attributable to indirect effects. There are however multiple ways in which hits can be attributed to indirect effects, depending on the balance and direction of indirect and direct effects. This is something we analyzed further in Sup. Figure 29.

1) If a minority of the phenotype is attributable to indirect effects, correction for this minor effect can still lower the ranking of a gene to below the hit threshold, and push other genes above the hit threshold. This is the most common case for the hit list changes shown in Fig. 4A, and occurs in all large-scale screens as indicated in Sup. Fig. 29.
2) Sometimes the majority of the phenotype can be attributable to indirect effects. Sup. Figure 29 shows that this occurs to a large extent in 3 of the 7 large-scale screens.
3) Finally, masked effects do occur in the large-scale screens (Sup. Fig. 29). However, it is rare that direct and indirect effects (strong enough to completely obscure a hit phenotype) fully cancel each other out (Sup. Fig. 29). However, as shown in point 1, mild masked effects do have strong influence on hit lists of all 7 DG screens.

Can this be corrected primarily by accounting for cell number since kinases will impact the cell number parameter substantially?

This cannot be done since correcting for cell number effects alone does not work because there can be changes in population context without changes in cell number (this was already shown in Sup. Fig. 14), and knowing changes in cell number alone cannot predict all possible changes in the

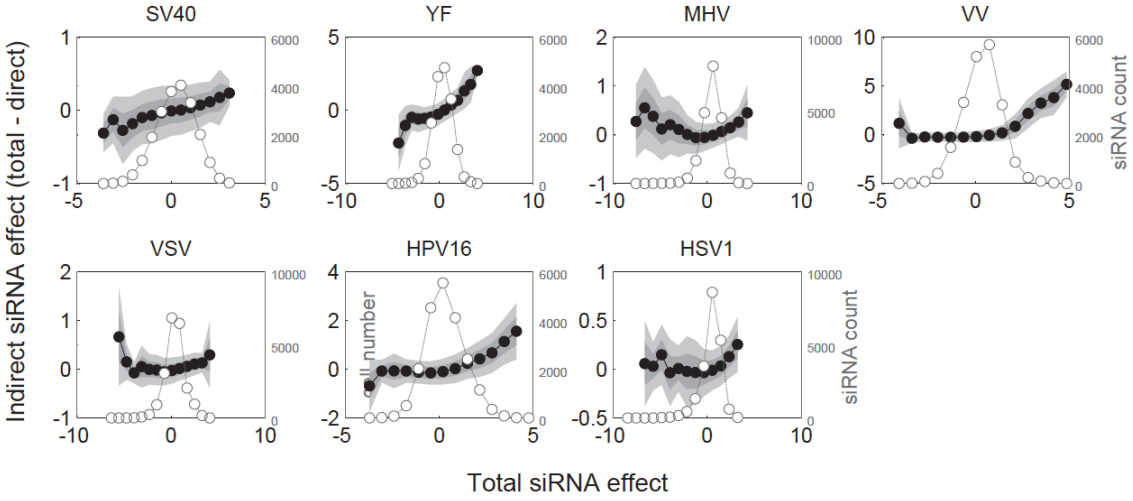
population context (see above and also below). It furthermore can lead to overfitting (see below). We have further tested if kinases affect cell numbers stronger compared to all other genes tested in the 7 druggable-genome screens (including phosphatases, transcription factors, ribosomal genes, proteasome genes, etc.), and see no evidence for the reviewer's assumption that cell number effects are particular for kinases.



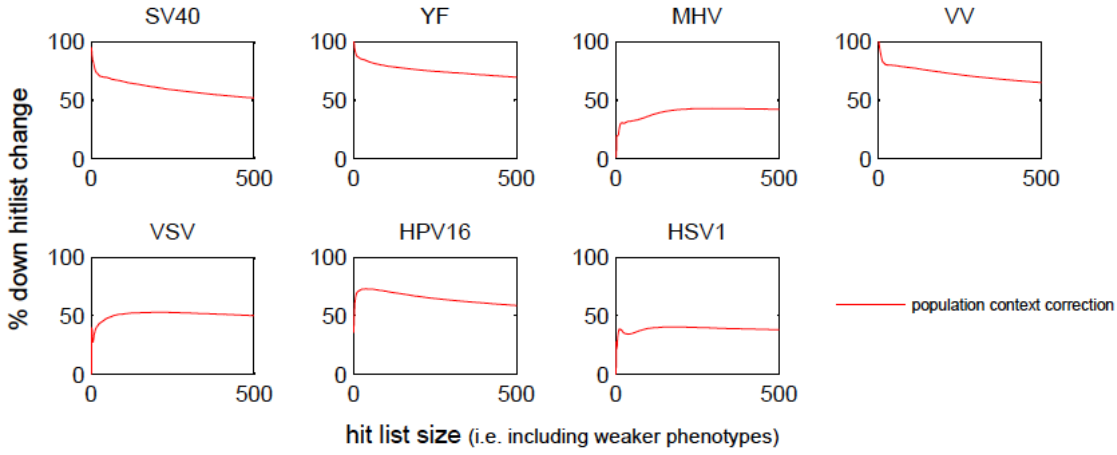
We have further characterized how simpler analyses (a total cell number correction using LOWESS, and a cutoff based method, both population-averaged and univariate methods) compare to our single-cell multivariate correction method in Sup. Figure 15. Lowess (the more advanced method of the two) only corrects 81 of the 203 genes in the same way as our method (for virus infection, on average over all 7 druggable genome screens), and introduces genes as a result of overfitting – i.e. without having a biologically motivated reason. As expected, the cutoff method performs even worse.

How strong are these effects versus direct effects? Can we preferentially filter a significant percentage of indirect effects by looking at their relative strength compared to direct effects? i.e. increase the cut-off for hits selection.

To answer the first question, we have compared indirect effect strength as a function of total effect strength for all 7 druggable genome screens, see below (showing mean and standard deviations of indirect effects aggregated over measurements with roughly similar total effects). This reveals that the strength and direction of indirect effects is very much assay dependent and cannot be known *a priori* by knowing the strength of the total effect.



We have further characterized how hit list change behaves as a function of hit list threshold (i.e. hit list size). This reveals that for some assays (including VACV, SV40, and YFV infection) indirect effects are not filtered out more effectively by taking more stringent hit lists (as evidenced by the higher % hit list change for smaller more stringent hit lists), as suggested by the reviewer.
In the cases where this would work (VSV, MHV, HSV1, HPV16), the advantage is only present for very stringent (small) hit lists (< 30 hits).



The authors mention both the growth-promoting first siRNA for DYRK3 and the growth-inhibitory effects of AK5, as both working indirectly via alterations in cell number, doesn't this strongly suggest that cell number alone can be used as a filter for indirect effects.

To what extent cell number corrections alone correct for population context effects (approx. 50%) was already shown in Sup. Fig. 14.

Figure 2D already showed how siRNA DYRK3_3 not just leads to a reduced total cell number, but also to a reduced local cell density (beyond what is explained by the total cell number change). This is consistent with the Bayesian analysis shown in Figure 3A for siRNA DYRK3_3 that inferred causal edges to both total cell number and local cell density, and explains why a total cell number correction alone is insufficient to account for population context effects.

Why use a complex analysis program that likely the majority of screeners will not be able to master if most of the effect can be controlled for or spotted using simpler approaches?

Correcting for cell number alone accounts for 50% of indirect effects. 50% in our opinion does not constitute "most of the effect". Furthermore, one cannot *a priori* know in which case only correcting for cell number effects would work without measuring the full population context. Therefore a cell number correction alone has the serious risk of overfitting/overcorrecting, as observed for the LOWESS correction.

To further address the reviewer's point we have implemented our correction method in CellProfiler, as described above, that makes it easy for everyone doing image-based screening to apply our methods. Many screeners have successfully adopted CellProfiler, so we are confident the same will be the case with our novel methods.

Finally, an important reason why people should use our methods is that it allows you to interpret and understand in more detail what is happening in your experiment. It is highly informative to understand the population context dependencies of your activity of interest. We feel that this should be basic knowledge for any molecular cell biologist, one that is currently missing or ignored.

The authors should please address this very practical implementation concern.

(See also above regarding our implementation in CellProfiler)

We agree with the reviewer and have therefore created a user-friendlier implementation of our novel methods. Specifically, we have created two modules (MeasurePopulationContext and PopulationContextCorrect) for CellProfiler (v1) that implement both our methods for the measurement of all features of the population context used in this study, as well as for the automated modeling and correction of the influence of the population context on any given cellular measurements. These modules are provided for download from the www.infectome.org website

with an example dataset and pipeline, and extensive documentation (on the website, and in the CellProfiler help function) showing how to use these modules.

Page 7 line 15 Figure S30 is mentioned as providing validation of direct effects. The effects shown with many of the siRNAs are moderate, in some cases less than two fold change. Therefore corresponding expression levels of the targeted proteins or mRNAs are needed to help accurately interpret if these effects are on target, either westerns or q-PCR to show that the level of knockdown matches the phenotype.

To address this concern we have performed q-PCR on the 2 DYRK3 siRNAs (DYRK3_2 and DYRK3_3) that gave strong phenotypes for 3 viruses (VSV, HSV, YFV), in 2 different HeLa strains (MZ & KY), in the original small-scale screens. Indeed, knockdown was respectively 83% and 99% efficient. DYRK3 was furthermore validated using 2 different small compound inhibitors as well as 2 additional siRNAs from a different vendor.

Additionally data should be shown for both the shRNA-mediated knockdown and the restored expression of AAK1. It is also problematic to rescue a knockdown by simply overexpressing a cDNA; it must be resistant to the expressed shRNA to demonstrate specificity. Is the shRNA against the 3'UTR of AAK1? Adequate information for this figure should be provided so that the results can be interpreted and reproduced, i.e. siRNA and shRNA sequences and methods. Also what do the error bars represent? How many times was this experiment performed?

We have added the information to the Supplementary Figure 31 and to the Materials and Methods section, and we apologize for the shortcoming. Indeed, the overexpressed cDNA was resistant to the shRNA, as the shRNA targeted the 3' UTR of AAK1, which was absent in the construct. Error bars represent means and standard deviations of 3 independent experiments.

Page 8 lines 13-17, the authors state that for the minority of the large screens, 3/7, masked or indirect effects accounted for 25-50% of all siRNA phenotypes and refer to Fig. 28 which also shows that for 4/7 screens the impact of masked and indirect was negligible.

Hit list change is more than 30% for all screens and therefore indirect effects are definitely not negligible. As discussed above, hit list change is not the same as amount of hits **completely** attributable to indirect effects. Accounting for even mild indirect effects can change the conclusion which genes give the strongest direct phenotypes and therefore warrant follow-up. We have rewritten the manuscript to avoid this confusion.

The authors should be less biased and more forthcoming about what the majority of their data shows regarding their hypotheses.

While we felt that we were very forthcoming, we believe that this statement is based on some of the misunderstanding, which we have addressed above. We have further added the analysis of 2 genome-wide screens.

Why is there such a stark difference in how the viruses behave in regard to the cell context?

That depends on the molecular biology underlying the cell population context dependencies (i.e. which molecules does the virus bind to on the cell, what signaling is required/induced for infection, which endocytic pathway does the virus require, etc.), as we have investigated and discussed in Snijder et al., Nature, 2009.

The SFV infection pattern-switchers Trio and Abl (and the other identified regulators of population context-dependent cell-to-cell variability) further corroborate and expand this interpretation.

It is hard with the garbled figure legends to tell what virus is shown for what graph, this should be corrected of course.

Naturally, we apologize for the problem that arose during the conversion of our Supplementary Information PDF by the Manuscript Tracking System. We would like to emphasize though that the original PDF file did not suffer the same problem.

However it suggests that cell context is NOT important for the majority of the larger screens that the authors have completed,

As discussed above, the hit list change (and subsequent improvements in screen and siRNA results) argues that cell context is important for all large-scale screens investigated. We have further supported this claim by the additional analysis of 2 genome-wide screens.

suggesting that it may be indirect modulations of discrete host cells pathways, i.e. endocytosis, that impact certain viruses more than others, or certain cell lines more than others. Some viruses use more than one host cells endocytic pathway to infect, does reliance on CME result in a more significant cell context dependent phenotype?

Reliance on CME does not appear to result in more significant cell context dependent phenotypes. The two strongest observed dependencies are SV40 and VSV infection in HeLa CNX in the small-scale screens (see Sup. Fig. 20). SV40 is not dependent on CME, while VSV is. Furthermore, SFV is a classic CME-dependent virus, and does not show a particularly strong population context dependency in unperturbed cases, as shown for instance in Figure 1D.

Again this argues that much of what the authors are discussing may be attributable to the cell context modulation of a few cellular actions, i.e. lipid regulation and CME for example, two pathways that were previously reported by the authors as being modulated by cell context.

While it is true that currently only a few cellular activities have been shown to depend on the cell population context, the presence of strong population context dependencies in 33 of the 34 tested infection assays argues that it is wide-spread. Especially since these 17 different viruses are known to have different infection pathways and host dependencies. Additional pathways shown (in Snijder et al, Nature, 2009) to be population context-dependent include for instance the regulation of cell size, cell adhesion signaling, and can also be expected to include cell-cell adhesion, migration, polarization, and other cellular processes. This is something we discuss in the introduction in this manuscript, as well as in Snijder & Pelkmans, NRMCB, 2011.

Therefore these large data sets suggest that if the phenotypic readout for a screen involves a cell context sensitive pathway then you should plot infection (or whatever your readout of interest is) against cell number. If there is a correlation observed then you might filter out "higher cell number" hits as a correction, or test the hits with different densities of cells per well in follow up assays.

We agree that testing for a correlation between the readout of interest and total cell number can be a first indicator of population context dependencies, however, absence of a correlation does not rule out complex context dependencies. As we have discussed above several times, only correcting for cell number effects is not sufficient and leads to overfitting. And, just filtering out “higher cell number” hits would be insufficient, as lower cell number hits can have equally strong or stronger population context-dependent effects (see Fig. 2).

Also since the kinase screens may have had more indirect effects per the authors; it suggests that they are not representative of the actions of siRNAs from a broader sampling of genes. This should be commented upon.

In the manuscript we did not suggest that silencing of kinases should have more indirect effects than that of other genes. For instance, as we have shown above, silencing of kinases does not overall lead to stronger cell number perturbations than silencing the rest of the druggable genome (see figure above). Furthermore, percentage hit-list change is roughly equal for the small-scale kinase screens and the large-scale druggable-genome screens and the genome wide screens (see Fig. 4A (was 5A)).

Did the same viruses that had virtually no indirect effects in the larger screens demonstrate a different level of susceptibility to indirect effects in their respective kinase screens? And did this correlate with cell numbers?

We realize that the labeling of Fig. S28 (now relabeled to Fig. S29) was a bit confusing, and we have changed it accordingly. The figure was meant to show siRNAs where the original measurement

(total effect) is fully masked or fully indirect. This does not mean the absence of indirect effect for those siRNAs that were labeled as direct in the original scatter plots. Any deviation from the diagonal indicates an indirect effect, also for the siRNAs previously labeled in green as "direct". The influence of these indirect effects on hit lists shows that indirect effects are present in all 7 druggable genome screens. For instance, in the MHV druggable genome screen, around one third of the hits changed (Fig 4A), although very few fully indirect or fully masked siRNA phenotypes were present (Fig S29).

There is no strong relationship between the amount of indirect effects measured in the 7 druggable genome screens and that in their corresponding small-scale screens. Clearly this depends also on the tested genes and their influence on the population context.

There is however a positive correlation between the strength of the population context-determined infection pattern and the amount of variability that we can attribute to indirect effects (as shown for the small-scale screens in Sup. Fig. 13).

The authors must publish or make available the percent infection data for ALL of the screens analyzed in this work, i.e. cell numbers, so that their results can be assesses appropriately by others.

We agree with the reviewer that it is important for the field of RNAi screening to make all data available, something that has thus far not been always the practice. For this manuscript, we have created a website that makes available all data relevant for a full assessment of the novel methods and results. This includes the following:

- All 34 small-scale RNAi screens of the 17 virus infections will be published in this study at www.infectome.org. This includes images (originally 800GB of image data) and full relevant measurements, that can be browsed and are searchable on the website.
- Entire SV40 infection data of the druggable genome screen will be published with this study as excel file. Image data will follow on the website, but we cannot guarantee its availability for this resubmission due to the technical challenges associated with the publication of such a large dataset.
- Entire cell size readouts and full population context parameters of all 7 druggable genome screens will be published with this study as excel file.

However, the infection results of 6 of the 7 large-scale screens, the cholesterol screens, and the endocytosis screens, will be presented in separate manuscripts that follow up on the new biology discovered. Except for the cholesterol screens, these datasets will then be released on the same (or related) website in exactly the same format.

They should also include the relevant imaging analyses parameters they have calculated

We agree with the reviewer, see above.

, so that the conclusions and methods can be assessed. This is very important since they are not focusing on specific siRNA hits but instead are strongly emphasizing the broad applicability of their methods and conclusions to the overarching data analysis and hit selection of entire screens to be used as a guiding resource for the RNAi screening community.

We agree with the reviewer, see above.

Why haven't cell context dependent events been observed or reported in image based small molecule screens for anti-viral compounds?

We can of course only guess. We think that the same explanation can be given as for why this hasn't yet been systematically characterized in image-based siRNA screens: Everyone has thus far been averaging single-cell readouts. However, we expect that a link between small molecule effects and cell population context will be reported in the future. For instance, in small compound screens heterogeneity in the response to drugs has been reported, although that has not been linked to population context effects. The Altschuler laboratory, for instance, studied how individual cancer cells phenotypically respond in highly variable ways to anti-cancer drugs. It is furthermore believed

that such non-genetic variability in these responses to drugs adds to the drug resistance of tumors (Brock et al., 2009, Nature Review Genetics). We have further discussed this in Snijder & Pelkmans, 2011, NRMCB.

One might imagine that the time course of the screen and the larger effects used for hit selection may have served to minimize the impact of compounds that increase or decrease cell growth. The authors should please comment on this […]

We have added a comment on this in the discussion section of the manuscript. While short drug treatments will less likely alter the cell population context as much as 3-day long siRNA treatments, it is possible that a drug target is expressed in a population context-dependent manner. The population averaging of small compound effects that inhibit targets with different expression patterns can thus be imagined to have an influence on the ranking of these compounds. It will be interesting to see the effect of population context normalization on small molecule screening results. Additionally, partially penetrant and switcher drug-phenotypes would be overlooked in population-averaged small molecule screens.

and also provide cell number data for their validation experiments using GSK-626616 since this is not provided in Fig 30 panels g and h.

We have added the requested cell number results to Fig S31 (was Fig S30).
Cell numbers relative to DMSO levels were on average over triplicates 54% for 1μM GSK-626616 and 71% for the 1μM Harmine treatments in the VSV_MZ validation experiment. Given the population context dependency of VSV infection in HeLa MZ cells (Fig. 1b), this would mean unperturbed infection levels at these reduced cell densities should be higher (as VSV infection is higher in sparse & edge HeLa MZ cells). Therefore, the **direct** drug effects inferred for both small compounds are actually stronger than presented in Sup. Fig. 31.
For the 1μM GSK-626616 treatments of HSV1 and YFV infection relative cell numbers were (similarly) 55% and 58% respectively. As both viruses closely resemble the infection patterns of VSV_MZ (see Fig. 1B), in both cases the direct effect of 1μM GSK-626616 presented in Sup. Fig. 31 will be even stronger.

In fact, the above represents a nice example of how an indirect effect of a small compound partially masks the direct effect of the compound. It doesn't fully mask the direct effect, but in the case of a large-scale screen, this might significantly affect its ranking.

Page 10 line 16, the authors should note that the 7% overlap is of exact genes, there was significantly more overlap in pathways and complexes as the Bushman paper made more clear.

We have included a section on the effects of our methods on functional annotation overlap between different screens. We have quantified the differences in enriched functional annotation overlap for the regulators of cell size between the 7 druggable genome screens (See Fig. 30). As expected by the increased exact gene overlap, we find that functional annotation overlap also improves considerably between the hit lists of the 7 screens. In contrast to the Bushman paper, we show exactly how many functional annotations are shared between how many assays and at what P-value threshold, for both total and direct effects.
In conclusion, and consistent with the improvement in gene hit overlap, not only does our method improve the exact gene overlap, it also increases the consistency observed for enriched functional annotations between hit lists.

In those instances different cell lines, viral preps, protocols and siRNA libraries were used. In the case of this work all of those were identical with the exception of the cell lines, which vary. Therefore the relevance of these observations to the comparison of screens between groups using much different approaches may be of less value. The authors should acknowledge the overlap in pathways and complexes in viral screen meta-analyses, as well as to the much greater homogeneity in their current screening efforts than those done by distinct groups.

The reviewer was correct in suggesting that our screens so far were all performed in the same experimental setup. Therefore, to test the improvements observed for our methods between screens from different setups (siRNA libraries, different cell lines, performed in different cities, with

different screening facilities and transfection protocols, by different people), we have further analyzed two genome-wide screens from Geneva/Lausanne.
To compare the effects of our analysis on hit list overlap between such heterogeneous screens we have calculated the overlap between the cell size results of our 7 druggable genome screens against each of the additional screens. On average we find 15% to 40% improved overlap between these hit lists. This now shows that also screens of a much more heterogeneous origin can be improved upon by our method, underscoring the relevance of our approach also for the comparison of screens from different laboratories using different approaches.

Page 10 bottom paragraph and figure 5d; it appears that only three of 15 screens improved in assay correlation with correction, suggesting that cell context variability occurs in the minority of screens presented.

This is a wrong interpretation of our results. As discussed above, correlation is a similarity measure that works on the measurements of the whole screen, including all the non-hits. Furthermore, those screens that did not improve upon correction in the small-scale screens showed already high (see Fig. 4D (was 5D)) correlations between results obtained on different HeLa strains before correction. To further clarify the difference between screen correlation and hit list overlap we have now calculated hit list overlap between those 34 small-scale screens performed with the same virus and we see improvements for 9 of the 15 viruses (while 4 stay the same and 2 decrease). On average, down-hit overlap improves from 26.7% → 40%, and up hits improve from 11.1% → 20.0%.

Are these the same viruses that showed the most indirect effects in Fig S28? It is hard to tell because of the garbled legends. Regardless, the authors should please comment on the general relevance of their correction if the majority of their screens did not improve in correlation after correction.

We have now clarified the difference between correlation measures and hit list overlap measures in the manuscript, and (as discussed above) have further analyzed the hit list overlap among those 34 small-scale screens performed with the same virus.

The viruses that increased most in their screen correlations upon correction are those viruses that showed significantly different patterns of infection depending on the infected HeLa strain (see Fig. 1B & D), as discussed in the main text. These are VACV, HSV1 and DEN-1. The viruses that showed most fully indirect and fully masked effects in the druggable genome screens shown in Fig. S29 (was Fig. S28) are SV40, YFV, VACV and HPV16. Only VACV is shared between these two lists.

Note that the first list (from Fig 4D (was 5D)) shows viruses for which the correlations improve because the pattern of infection was different depending on the HeLa strain, while the second analysis (Fig S29) shows druggable genome screens that displayed **fully** indirect or **fully** masked effects. This is not the same. For instance, we see small-scale screens with strong indirect effects (such as SV40_CNX, as shown in Fig. 1G, 2F), for which both the total and direct effects nonetheless correlate very well between different HeLa strains (Fig 4D).

2nd Editorial Decision 28 February 2012

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the referees who accepted to evaluate the revised manuscript. As you will see, the referees find the topic of your study of potential interest and are now globally supportive.

They raise however a series of remaining minor concerns, which we would ask you to carefully address with suitable amendments to the manuscript.

On a more editorial note, we would also kindly ask you to address the following points:

- We tend to agree with referee #1 that the expression "deep screening" might not be well defined

and the title should be amended accordingly to better reflect the content of this study.

- On the infectome.org platform, it appears that the datasets include only summary values (median and deviation) for (some of) the features measured. Perhaps we have misunderstood how to use this resource, but we could not find a way to download the full dataset of extracted features at the single cell level. The manuscript refers to the measurements of "200 quantitative features of 2.4x10^9 cells". The complete dataset of single-cell measurements across the 34 small-scale screens appears to be central and integral to the present study since this full dataset is necessary to enable reproduction and testing of the single-cell modeling and phenotypic classifications reported here (see an example of download page here: http://www.ebi.ac.uk/huber-srv/cellmorph/data.php related to the article http://www.nature.com/msb/journal/v6/n1/full/msb201025.html). The same issue extends to the SV40 infection druggable genome screen.

- With regard to access to the images, we appreciate that the images are made available on infectome.org on a gene-by-gene basis. However, for researchers interested in a systematical re-analysis of these images, for example to train or test feature extraction algorithms, bulk download of the full set of annotated images would be much more useful (see again http://www.ebi.ac.uk/huber-srv/cellmorph/data.php).

- To provide attribution to the authors who performed the unpublished "Lausanne" screens, we would recommend to include a citation in the main text in the form of "(Author et al, in preparation)".

- Password protection of interactome.org should be removed upon publication of this paper.

Thank you for submitting this paper to Molecular Systems Biology.

Yours sincerely,

Editor
Molecular Systems Biology

http://www.nature.com/msb

Referee reports

Reviewer #1 (Remarks to the Author):

Overall I find this version a marked improvement on the earlier one, and remain extremely supportive of this interesting work. Unfortunately I still feel some revisions are necessary in order to still further remove statements that are qualitative/biased interpretations of their own analysis. The work is impressive on its own right, and the authors shouldn't feel the need to overstate their claims. I do not feel this manuscript should be published until some of these lingering issues are finally addressed.

Specifically,

1) Section on "biological interpretation".

"Accounting for population context-mediated effects most strongly removed genes from the hit list that were annotated with alternative splicing, and genes associated with the extracellular space (Figure 5, and Supplementary Figure 27). Accounting for population context-propagated effects therefore not only improves the reproducibility and consistency of RNAi screening results at the siRNA, gene, and functional annotation level, but also improves the biological interpretation of those results."

My biggest issue concerns this section, which was not in my opinion appropriately revised from the

previous version.

"Biological interpretation" is not to my knowledge a metric that can be quantified, thus I am very unsure how it can be improved upon. As in the previous version of the manuscript, there seems to be some implication that the because there is decreased enrichment in splicing factors and genes associated with the extracellular regions, and an increase in other functional categories which the authors assume should be enriched, that somehow the data from the screens has been improved. Besides the fact that it is possible these particular categories went up/down simply due to chance, why this enrichment is necessarily more "biologically interpretable" is very unclear to me. Unless the authors are prepared to explain this statement at greater length, I would suggest this section be reworked.

An alternative way to describe this is that functional categories that are decreased after taking into account cell context are those that contain genes that are more likely to result in indirect effects. Such effects could be result from the inhibition cellular functions (such as splicing) that have pleiotropic roles in regulating numerous aspects of cellular physiology (that are either genuine or potential off-target) that not only confound specific phenotypic readout such as virus entry, but are often are isolated in many other types of screens.

2) Other statements

"In an era where large-scale RNAi screens are increasingly performed to reach a systems-level understanding of cellular processes, we show that this is best achieved by analyses that account for and incorporate the single-cell microenvironment."

I think "best achieved" seems like a strong statement here. Although I am convinced that accounting for population context has significant effect in many cases, whether this is truly generalizable cannot be assessed from this work. I do not think they have shown (yet) that accounting for context is always necessary.

"In conclusion, hit lists from RNAi screens are strongly confounded by indirect and masked effects due to changes in the cell population context. Correcting for these unwanted effects strongly changes such lists, improves the statistical window for hit scoring, and increases reproducibility between screens performed in different cell line strains."

Again.... I do not think "strongly confounded", and "strongly changes" are warranted. The authors observe a genuine effect in some of the screens they analyzed, not an overwhelming effect in all screens.

"RNAi in human tissue culture cells can lead to strong changes in population context parameters (Figure 2A-C)"

Do the authors mean that RNAi can affect cell and population phenotypes, such as size and local cell density, which can in turn affect virus entry? If so this is an example of where the authors make statements (and use somewhat over-complex figures) say something that is relatively intuitive and said in a much more straightforward manner. I would assume many siRNAs can change size/density etc..

"However, published RNAi screens of the same virus infection have had poor hit list overlap (between 3% to 6% overlap in hits from several HIV screens) (Bushman, Malani et al. 2009; Cherry 2009; Mohr, Bakal et al. 2010). Based on our findings so far, we reasoned that this could, in part, be caused by the fact that population context-determined cell-to-cell variability greatly depends on culture conditions and that viruses can display different population context-dependent patterns of infection in different host cell line strains (see Figure 1B)."

These are two disconnected issues. The poor overlap exists, but whether this has anything to do with population connect still truly remains to be investigated.

"If population context-mediated effects should generally be accounted for is currently not clear, but we expect the cell population context to influence the expression of a large number of genes (see for

instance (Fukuhara, Sako et al. 2008)), metabolic activity, signal transduction, cell polarization, cell proliferation and migration, and the determination of cell shape (Kim, Lee et al. 2010; Schultz, Marenstein et al. 2011; Snijder and Pelkmans 2011)."

Given this sentence is in the discussion, it is probably not a major issue. But I think there is a long way to go before I am convinced that population context will affect the outcomes of all these screens.

3) "Deep screening".

I personally question whether the introduction of this term is necessary here and/or appropriate. For one, the real value of this manuscript is the analysis of context-effects across multiple screens, and not simply the fact they have performed many screens - thus the title is misleading. Secondly, if deep screening were the same as deep sequencing, wouldn't that mean use of more/different siRNAs in the same cell line?

Minor comment.

In a case like DYRK (where 2 siRNAs give one phenotype, and 1 siRNA gives another), a statement in the main text as to how this inconsistency is dealt with prior to the determination of final gene hits is warranted. I assume the inconsistent siRNA is eliminated and the final score is the result of the 2 consistent siRNA. One sentence here would be helpful.

Reviewer #2 (Remarks to the Author):

In this revised version of their manuscript the authors have addressed the concerns raised initially. Their efforts to improve, explain and make more accesible their work is admirable and will hopefully add to the usefulness of this innovative approach for analyzing image-based screens and experiments.

A major advantage of their using the Geneva/Lausanne data sets is that that work was performed using a different siRNA library. This is important because improving the overlap between orthologous sets of siRNA reagents will also improve the strength of using reagent redundancy thereby minimizing false negatives and false positives, something which must be carefully considered and controlled for even with the addition of this new approach.

Comments:
I am hopeful that the two downloadable analysis modules will be sufficiently easy to use so that image-based screeners can readily take advantage of the Snijder-Pelkmans correction. Therefore I would ask that the authors please implement some way of modulating and/or explaining any common issues brought up by users over time.

2nd Revision - authors' response 09 March 2012

Response to editor:

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the referees who accepted to evaluate the revised manuscript. As you will see, the referees find the topic of your study of potential interest and are now globally supportive.
They raise however a series of remaining minor concerns, which we would ask you to carefully address with suitable amendments to the manuscript.

On a more editorial note, we would also kindly ask you to address the following points:

- We tend to agree with referee #1 that the expression "deep screening" might not be well defined and the title should be amended accordingly to better reflect the content of this study.

We now suggest: “Single-cell analysis of population context advances RNAi screening at multiple levels”. We hope you agree this reflects better the content of this study, but we are open to other suggestions.

- On the infectome.org platform, it appears that the datasets include only summary values (median and deviation) for (some of) the features measured. Perhaps we have misunderstood how to use this resource, but we could not find a way to download the full dataset of extracted features at the single cell level. The manuscript refers to the measurements of "200 quantitative features of 2.4x10^9 cells". The complete dataset of single-cell measurements across the 34 small-scale screens appears to be central and integral to the present study since this full dataset is necessary to enable reproduction and testing of the single-cell modeling and phenotypic classifications reported here (see an example of download page here: http://www.ebi.ac.uk/huber-srv/cellmorph/data.php related to the article http://www.nature.com/msb/journal/v6/n1/full/msb201025.html). The same issue extends to the SV40 infection druggable genome screen.

Similar to the Huber & Boutros paper referenced, we now allow the bulk download of a total of ~1.5TB of screening data and analysis results, encompassing both the raw images and all extracted single-cell features, as well as image segmentation results, preview JPG files, and SVM classification results, for the complete set of 34 small-scale screens and for the entire druggable genome screen of SV40 infection. Instructions for downloading the complete datasets are given on www.infectome.org/download-data.html. The raw data itself is can be browsed and downloaded from www.infectome.ethz.ch/data. Note however that it will take another week for us to copy over all the data (i.e. by 15th of March the data will be complete).

- With regard to access to the images, we appreciate that the images are made available on infectome.org on a gene-by-gene basis. However, for researchers interested in a systematical re-analysis of these images, for example to train or test feature extraction algorithms, bulk download of the full set of annotated images would be much more useful (see again http://www.ebi.ac.uk/huber-srv/cellmorph/data.php).

We agree, see above.

- To provide attribution to the authors who performed the unpublished "Lausanne" screens, we would recommend to include a citation in the main text in the form of "(Author et al, in preparation)".

We have added an appropriate reference as suggested: “(Scott, Vossio et al. In preparation)”, page 11.

- Password protection of interactome.org should be removed upon publication of this paper.

Upon publication of the manuscript we will remove all password protections present on www.infectome.org.

Reviewer #1 (Remarks to the Author):

Overall I find this version a marked improvement on the earlier one, and remain extremely supportive of this interesting work. Unfortunately I still feel some revisions are necessary in order to still further remove statements that are qualitative/biased interpretations of their own analysis. The work is impressive on its own right, and the authors shouldn't feel the need to overstate their claims. I do not feel this manuscript should be published until some of these lingering issues are finally addressed.

Specifically,
1) Section on "biological interpretation".
"Accounting for population context-mediated effects most strongly removed genes from the hit list that were annotated with alternative splicing, and genes associated with the extracellular space (Figure 5, and Supplementary Figure 27). Accounting for population context-propagated effects therefore not only improves the reproducibility and consistency of RNAi screening results at the siRNA, gene, and functional annotation level, but also improves the biological interpretation of those results."
My biggest issue concerns this section, which was not in my opinion appropriately revised from the previous version.
"Biological interpretation" is not to my knowledge a metric that can be quantified, thus I am very unsure how it can be improved upon. As in the previous version of the manuscript, there seems to be some implication that the because there is decreased enrichment in splicing factors and genes associated with the extracellular regions, and an increase in other functional categories which the authors assume should be enriched, that somehow the data from the screens has been improved. Besides the fact that it is possible these particular categories went up/down simply due to chance, why this enrichment is necessarily more "biologically interpretable" is very unclear to me. Unless the authors are prepared to explain this statement at greater length, I would suggest this section be reworked.
An alternative way to describe this is that functional categories that are decreased after taking into account cell context are those that contain genes that are more likely to result in indirect effects. Such effects could be result from the inhibition cellular functions (such as splicing) that have pleiotropic roles in regulating numerous aspects of cellular physiology (that are either genuine or potential off-target) that not only confound specific phenotypic readout such as virus entry, but are often are isolated in many other types of screens.

We have changed “biological interpretation” with “functional annotation”, and we have changed the 2 sentences quoted above (which are on the bottom of page 13) to reflect the reviewers suggestion: “Accounting for population context-mediated effects most strongly reduced annotations with expected pleiotropic roles in numerous aspects of cellular physiology, such as alternative splicing (Figure 5, and Supplementary Figure 27). Accounting for population context-propagated effects therefore not only improves the reproducibility and consistency of RNAi screening results at the siRNA, gene, and functional annotation level, but also increases the focus of screening results on genes that are directly involved in the intracellular activity screened for.”

2) Other statements
"In an era where large-scale RNAi screens are increasingly performed to reach a systems-level understanding of cellular processes, we show that this is best achieved by analyses that account for and incorporate the single-cell microenvironment."
I think "best achieved" seems like a strong statement here. Although I am convinced that accounting for population context has significant effect in many cases, whether this is truly generalizable cannot be assessed from this work. I do not think they have shown (yet) that accounting for context is always necessary.

We have changed “best achieved” for “is often improved” in the abstract.

"In conclusion, hit lists from RNAi screens are strongly confounded by indirect and masked effects due to changes in the cell population context. Correcting for these unwanted effects strongly changes such lists, improves the statistical window for hit scoring, and increases reproducibility between screens performed in different cell line strains."
Again.... I do not think "strongly confounded", and "strongly changes" are warranted. The authors observe a genuine effect in some of the screens they analyzed, not an overwhelming effect in all screens.

We have removed both occurrences of the word “strongly” (halfway page 10).

"RNAi in human tissue culture cells can lead to strong changes in population context parameters (Figure 2A-C)"
Do the authors mean that RNAi can affect cell and population phenotypes, such as size and local cell density, which can in turn affect virus entry? If so this is an example of where the authors make

statements (and use somewhat over-complex figures) say something that is relatively intuitive and said in a much more straightforward manner. I would assume many siRNAs can change size/density etc..

We have changed the sentence on the bottom of page to: “RNAi in human tissue culture cells can strongly affect population context parameters, such as cell size and local cell density (Figure 2A-C).”

"However, published RNAi screens of the same virus infection have had poor hit list overlap (between 3% to 6% overlap in hits from several HIV screens) (Bushman, Malani et al. 2009; Cherry 2009; Mohr, Bakal et al. 2010). Based on our findings so far, we reasoned that this could, in part, be caused by the fact that population context-determined cell-to-cell variability greatly depends on culture conditions and that viruses can display different population context-dependent patterns of infection in different host cell line strains (see Figure 1B)."
These are two disconnected issues. The poor overlap exists, but whether this has anything to do with population connect still truly remains to be investigated.

We have made our phrasing more accurate and now say (on bottom of page 9): “Based on our findings so far, we reasoned that poor overlap in cell population-averaged RNAi screens could, in part, be caused by different patterns of cell-to-cell variability (see Figure 1B).”

"If population context-mediated effects should generally be accounted for is currently not clear, but we expect the cell population context to influence the expression of a large number of genes (see for instance (Fukuhara, Sako et al. 2008)), metabolic activity, signal transduction, cell polarization, cell proliferation and migration, and the determination of cell shape (Kim, Lee et al. 2010; Schultz, Marenstein et al. 2011; Snijder and Pelkmans 2011)."
Given this sentence is in the discussion, it is probably not a major issue. But I think there is a long way to go before I am convinced that population context will affect the outcomes of all these screens.

We agree that it will take time before all this is convincingly shown. Although not mentioned, we believe the impact of regulated cell-to-cell variability will become even more apparent once people factor in time-resolved aspects of the population context such as clonality, varying growth speeds, and cellular age.

3) "Deep screening".
I personally question whether the introduction of this term is necessary here and/or appropriate. For one, the real value of this manuscript is the analysis of context-effects across multiple screens, and not simply the fact they have performed many screens - thus the title is misleading. Secondly, if deep screening were the same as deep sequencing, wouldn't that mean use of more/different siRNAs in the same cell line?

We suggest changing the title to: “Single-cell analysis of population context advances RNAi screening at multiple levels” and removed references to deep screening from the introduction and discussion.

Minor comment.
In a case like DYRK (where 2 siRNAs give one phenotype, and 1 siRNA gives another), a statement in the main text as to how this inconsistency is dealt with prior to the determination of final gene hits is warranted. I assume the inconsistent siRNA is eliminated and the final score is the result of the 2 consistent siRNA. One sentence here would be helpful.

We have added the following sentence to the bottom of page 8 (continuing on page 9): “To account for inconsistent siRNA phenotypes we calculated the median value of all three siRNAs, which effectively selects away any outlier siRNA phenotype.”

Reviewer #2 (Remarks to the Author):
In this revised version of their manuscript the authors have addressed the concerns raised initially. Their efforts to improve, explain and make more accesible their work is admirable and will

hopefully add to the usefulness of this innovative approach for analyzing image-based screens and experiments.
A major advantage of their using the Geneva/Lausanne data sets is that that work was performed using a different siRNA library. This is important because improving the overlap between orthologous sets of siRNA reagents will also improve the strength of using reagent redundancy thereby minimizing false negatives and false positives, something which must be carefully considered and controlled for even with the addition of this new approach.
Comments:
I am hopeful that the two downloadable analysis modules will be sufficiently easy to use so that image-based screeners can readily take advantage of the Snijder-Pelkmans correction. Therefore I would ask that the authors please implement some way of modulating and/or explaining any common issues brought up by users over time.

We have added a forum dedicated to answering software related questions on http://www.infectome.org/software.html & http://www.infectome.org/support-forum.html where people can post questions and report problems relating to the software released on the infectome website.

We have furthermore in the past been in contact with the CellProfiler team, who have previously offered to further support our custom modules. We will ask them to support these modules and where possible transform them to CellProfiler 2 compatible modules.

